United States
Environmental Protection
Agency

Research and Development

Atmospheric Research and
Exposure Assessment Laboratory
Research Triangle Park, NC 27711

EPA/600/SR-92/221 January 1993

EPA

Project Summary

Application of a Data-Assimilating Prognostic Meteorological Model to Two Urban Areas

Sharon G. Douglas

A data-assimilating prognostic meteorological model, the Systems Applications International Mesoscale Model (SAIMM), was applied to generate meteorological fields suitable for photochemical modeling of two urban areas: Los Angeles, California, and the Lower Lake Michigan area which includes Chicago, Illinois. The objectives of this study were to test the ability of the SAIMM to provide accurate meteorological fields for photochemical modeling of the Los Angeles and Lower Lake Michigan urban areas and to investigate the meteorological data requirements needed to support the use of the SAIMM four-dimensional data assimilation (FDDA) procedure.

Testing of the SAIMM/FDDA methodology was accomplished through a series of nudging-effectiveness and data-reduction simulations. For Los Angeles, the SAIMM/FDDA procedure was tested using observational data collected during the 1987 Southern California Air Quality Study (SCAQS) and was applied to 25 June (one of the SCAQS episode days); for the Lower Lake Michigan area the procedure was tested using observational data collected during the 1991 Lake Michigan Ozone Study (LMOS) and was applied to 26 June (one of the LMOS episode days). The results of the nudging-effectiveness experiments for both the Los Angeles and Lower Lake Michigan areas indicate that assimilation of both wind and temperature data provides the best representation of the meteorological fields. The data-reduction simulation results indicate that even when the intensive SCAQS and LMOS data sets are reduced to what might be considered routine data sets, assimilation of the available wind and temperature data provides an improved representation of the meteorological fields when compared with a simulation in which no data assimilation was performed. Appropriate specification of the episode and domain-dependent analysis and modeling parameters is essential to successful application of the technique.

This Project Summary was developed by EPA's Atmospheric Research and Exposure Assessment Laboratory, Research Triangle Park, NC, to announce key findings of the research project that is fully documented in a separate report of the same title (see Project Report ordering information at back).

Introduction

In this study we have used a data-assimilating prognostic meteorological model, the Systems Applications International Mesoscale Model (SAIMM), to generate meteorological fields suitable for photochemical modeling of two urban areas: Los Angeles, California, and the Lower Lake Michigan area which includes Chicago, Illinois; Milwaukee, Wisconsin; Gary, Indiana; and Muskegon, Michigan. These areas were selected for study because both of the areas (1) have been designated ozone non-attainment areas by the U.S. Environmental Protection Agency (EPA) and continue to experience exceedances of the National Ambient Air Quality Standard (NAAQS) for ozone, (2) are characterized by complex mesoscale meteorology that cannot be accurately represented by routinely collected meteorological data, and (3) have been the setting for recent intensive air quality/meteorological data collection studies, and thus enhanced surface and upper-air meteorological data are available for a number of ozone episode days.

The objectives of this study were to test the ability of the SAIMM to provide accurate meteorological fields for the Los Angeles and Lower Lake Michigan urban areas and to investigate the meteorological data requirements needed to support the use of the SAIMM four-dimensional data assimilation (FDDA) procedure. To this end, a series of nudging-effectiveness and data-reduction experiments were performed for each of the areas. Model performance for Los Angeles was evaluated (both graphically and statistically) using data from the 1987 Southern California Air Quality Study (SCAQS); for the Lower Lake Michigan area, data from the 1991 Lake Michigan Ozone Study (LMOS) were used.

Testing and Evaluation Procedures

Modeling Procedures

Prognostic meteorological models numerically solve an approximation of the equations that govern atmospheric behavior. Beginning with a set of initial conditions that represent the state of the atmosphere at the initial time, a prognostic model simulates the response of the atmosphere within the domain of interest to differential heating of the earth's surface. Prognostic models provide a dynamically consistent, physically realistic, three-dimensional representation of the wind as well as other meteorological variables, such as potential temperature, and planetary-boundary-layer height. However, the prognostic model solution may not always replicate observations and, therefore, may not accurately represent day-specific meteorology as is required if the fields are to be used for air quality modeling. Numerical approximations, physical parameterizations, and initialization problems represent a few of the potential sources of error in meteorological models that can cause the model solution to deviate from actual atmospheric behavior.

The objective of four-dimensional data assimilation (FDDA) is to improve the agreement between the simulated fields and observed data and, thus, provide more accurate meteorological fields for photochemical modeling of historical episode days. Using this procedure, observed data are incorporated into the prognostic model solution during the course of a simulation. The most common approach to FDDA is Newtonian "nudging" in which the prognostic variables are relaxed or "nudged" toward the observational data by additional forcing terms in the prognostic model equations. The general form of the prognostic equation for a variable a is

$$\frac{\partial a}{\partial t} = F(a,\underline{x},t) + G \times w_t \times w_{xy} \times (a' - a)$$

The first term on the right-hand side of equation (1), F, represents all of the model's physical processes. The second term is the nudging term. G determines the relative weight of the nudging term with respect to the model's physical processes. Typical values of G are 10^{-3} s^{-1} for strong nudging and 10^{-4} s^{-1} for weak nudging. The variables w_t and w_{xy} are temporal and spatial weighting functions and the quantity a' represents the analyzed or observed value.

Data assimilation is accomplished in the SAIMM using the Newtonian nudging technique. An objective analysis of the observational data is performed and a spatial weighting factor (based on data availability) is calculated. A temporal weighting factor varies linearly from 0 to 1 throughout the data assimilation interval. The degree to which the prognostic variables are then nudged toward the objective analysis is determined by the weighting information.

Overview of the Numerical Experiments

The numerical experiments were designed to test the ability of the SAIMM/FDDA procedure to provide accurate meteorological fields for the Los Angeles and Lower Lake Michigan urban areas and to investigate the meteorological data requirements needed to support the use of the SAIMM/FDDA methodology in each of these areas. For Los Angeles, the SAIMM/FDDA procedure was used to simulate the meteorology of 25 June 1987 (one of the SCAQS episode days). For the Lower Lake Michigan area, the simulations were focused on 26 June 1991 (one of the LMOS episode days).

The first series of numerical simulation experiments, referred to as the nudging-effectiveness simulations, were designed to examine the effectiveness of nudging the prognostic wind and temperature variables separately and in combination with one another and to determine (roughly) the optimum nudging coefficients for each.

Use of the intensive data from SCAQS and LMOS allowed us, in a second series of experiments, to investigate (through stepwise reduction of the input data), the data requirements for successful FDDA. For the data-reduction experiments, the input data base was reduced in three stages so that after the third reduction the data base approximated that which would be available from a routine monitoring network. For these experiments, the spatial density of the monitoring sites was reduced, but the temporal distribution of the observations at each site was not changed. A model run was performed with the full data set, and then again after each site reduction to determine how the model adjusted to a decrease in the spatial density of the observational data.

Evaluation Measures

A number of graphical and statistical analysis products were used to evaluate the simulation results. Graphical analysis was used to subjectively assess how well the assimilated data were represented in the meteorological fields and the effect of the data on the simulated fields in areas removed from the monitoring locations. Graphical analysis products include x-y, x-z, y-z, and z-t cross sections of the wind and temperature fields for several simulation times and locations.

Statistical analysis was used to quantify the differences between the simulated fields and the observed data and, thus, provided a basis for evaluating both the nudging-effectiveness and data-reduction experiments. The statistical analysis included the calculation of a number of statistical measures of bias including the mean residual, mean unsigned error, mean relative error (normalized bias), mean unsigned relative error (gross error), and the root mean square error.

Results

Testing and Evaluation for the Los Angeles Domain

The SAIMM/FDDA procedure was tested using observational data collected during the 1987 Southern California Air Quality Study (SCAQS) and was applied to 25 June (one of the SCAQS episode days).

Specification of the modeling domain (including the horizontal and vertical resolution) was based on geographical and meteorological considerations. The complex meteorology of this region is strongly influenced by the diurnal land/sea breeze cycle and by slope flows that develop along the steep terrain encompassing the Los Angeles basin. Therefore, the modeling domain includes the Los Angeles basin, adjacent offshore areas, and the surrounding terrain. The domain consists of 65 grid points in the west-east direction, 36 grid points in the south-north direction, and 22 vertical levels. The horizontal grid spacing is 5 km.

The simulation period used in this study included a full diurnal cycle, beginning and ending at 2300 LST. The SAIMM simulations were initialized using domain-scale profiles of temperature and specific humidity that were based on available meteorological sounding data from the Ontario, CA (ONT), monitoring site. The geostrophic wind, which is used to initialize the wind field and as a forcing term in the pressure gradient term of the momentum equations, was set equal to zero for all simulations. Because the assimilated data contain information on all scales of motion (including those too large to be resolved within the modeling domain), we have assumed that it is not necessary to artificially impose the large-scale forcing.

Nudging-effectiveness simulations in which wind data, temperature data, and wind and temperature data, respectively, were assimilated indicate that, for this SAIMM application, assimilation of wind data alone improves the agreement between the simulated and observed winds and the representation of the wind field but does little to improve the agreement between the simulated and observed upper-air temperatures. Assimilation of temperature data alone improves accuracy with which the upper-air temperatures are simulated but does little to improve the agreement between the simulated and observed winds. Assimilation of both the wind and temperature data not only improves the agreement between the simulated and observed winds and the agreement between the simulated and observed upper-air temperatures, but actually results in better agreement between the simulated and observed upper-air temperatures than does assimilation of temperature data alone.

The first two nudging-effectiveness simulations utilized nudging coefficients equal to 0.001 for wind and 0.0001 for temperature, respectively. The importance of the wind field in air-quality modeling and the indirect benefits derived from the assimilation of the wind data support the use of a larger nudging coefficient for the assimilation of the wind data than for the assimilation of the temperature data. However, further analysis of the simulation results indicated that strong nudging of the wind components toward the analyzed data created some unrealistic airflow patterns over regions where data were not available. Therefore, the nudging coefficient for assimilation of the wind data was reduced to 0.0005 in the third nudging-effectiveness simulation.

A series of data-reduction experiments, using the SCAQS data base, were performed to investigate the response of SAIMM/FDDA methodology to varying levels of data availability. To accomplish this, monitoring sites were eliminated from the data set in a series of three site-reduction exercises. A model run was performed with the full data set, and then again after each site reduction to determine how the model adjusted to decreased amounts of observational data.

Although necessarily somewhat subjective, the data reductions were based primarily on geographic considerations. The goal was to produce, after the third site reduction, a data set which represented a routine meteorological monitoring network. For the Los Angeles basin, which contains an extraordinary number of routine surface meteorological monitoring sites, this meant reducing the number of sites beyond what is normally available for this area. The number of surface and upper-air monitoring sites used for each of the data-reduction experiments is given in Table 1. The nudging coefficients were assigned based on the results of the nudging-effectiveness simulations and were set equal to 0.0005 for the u and v wind components and 0.0001 for temperature.

Table 1. SCAQS Site Reductions

Data Reduction	*Surface Wind Sites*	*Upper Wind Sites*	*Upper Temperature Sites*
0	*71*	*15*	*14*
1	*51*	*10*	*10*
2	*32*	*7*	*7*
3	*15*	*4*	*4*

A thorough graphical and statistical analysis of the data-reduction simulation results indicate that, even when the data set is reduced to what might be considered a routine data set, assimilation of the available wind and temperature data provides an improved representation of the meteorological fields when compared with the no-FDDA simulation. The influence of the data on the simulations is not confined to the monitoring site locations but is propagated within the modeling domain and influences the evolution of the meteorology over data-sparse areas as well.

Testing and Evaluation for the Lower Lake Michigan Domain

For the Lower Lake Michigan area, the SAIMM/FDDA procedure was tested using observational data collected during the 1991 Lake Michigan Ozone Study (LMOS) and was applied to 26 June (one of the LMOS episode days).

The Lower Lake Michigan area includes Chicago, Illinois; Milwaukee, Wisconsin; Gary, Indiana; and Muskegon, Michigan. The lake breeze (driven by the horizontal temperature gradients created by the differential heating of the land and water surfaces) plays an important role in determining the meteorology of this area and results in complex mesoscale circulation patterns along the lake shore. To allow resolution of the lake-induced circulations, the entire southern portion of Lake Michigan is included in the modeling domain. The domain consists of 50 grid points in the west-east direction, 52 grid points in the south-north direction, and 20 vertical levels. The horizontal grid spacing is 5 km.

The simulation period for the Lower Lake Michigan area simulations included a full diurnal cycle—extending from 2300 CST 25 June to 2300 CST 26 June. The SAIMM simulations were initialized using domain-scale profiles of temperature and specific humidity that were based on sounding data from the Kankakee, IL (KANK) monitoring site. As in the SCAQS simulations, the geostrophic wind was set equal to zero.

The results of the first three nudging-effectiveness simulations in which wind data, temperature data, and wind and temperature data, respectively, were assimilated were quite disappointing. While assimilation of the observed data improved the agreement between the simulated and observed winds and upper-air temperatures locally, some physically unrealistic meteorological features appeared in the simulated wind and temperature fields. Apparently, the information provided by the data had little effect on the simulation in data-sparse areas (i.e., this information was not propagated throughout the modeling domain). In preparing the analyses for FDDA, the user must specify maximum radii of influence for the interpolation of the data at the surface and aloft. In these initial simulations, the maximum radius of influence for the surface level was 20 km; aloft this value was set equal to 50 km. Additional objective analyses were prepared using a maximum radii of influence of 50 and 100 km for the surface and upper levels, respectively. An additional nudging-effectiveness simulation was performed using the revised analyses. Increasing the radii of influence in the objective analysis of the data constrained the simulation over a broader geographical area and contributed to much improved simulation results. Use of a larger radius of influence for the interpolation of data over the Lower Lake Michigan domain than for the Los Angeles domain is justifiable due to the absence of terrain in the Lake Michigan area. Nudging coefficients for this simulation were 0.0005 and 0.0001, respectively, for the wind and temperature variables. Assimilation of both the wind and temperature data improved the agreement between the simulated and observed winds and the agreement between the simulated and observed upper-air temperatures. As in the SCAQS nudging-effectiveness simulations, assimilation of the wind data in combination with the temperature data resulted in better agreement between the simulated and observed upper-air temperatures than assimilation of temperature data alone.

A series of data-reduction experiments, using the LMOS data base, were performed to investigate the response of SAIMM/FDDA methodology to varying levels of data availability. To accomplish this, monitoring sites were eliminated from the data set in a series of three site-reduction exercises. A model run was performed with the full data set, and then again after each site reduction to determine how the model adjusted to decreased amounts of observational data in a similar manner to the SCAQS runs.

The nudging coefficients were assigned based on the results of the nudging-effectiveness simulations and were set equal to 0.0005 for the u and v wind components and 0.0001 for temperature. The analyses for the data-reduction simulation were prepared using the larger radii of influence (50 km at the surface and 100 km aloft).

A thorough graphical and statistical analysis of the data-reduction simulation results indicates that even when the data set is reduced to what might be considered a routine data set, assimilation of the available wind and temperature data provides an improved representation of the meteorological fields when compared with the no-FDDA simulation.

Summary and Recommendations

Testing of the SAIMM/FDDA methodology for application to the Los Angeles and Lower Lake Michigan urban areas was accomplished through a series of nudging-effectiveness and data-reduction simulations. For Los Angeles the SAIMM/FDDA procedure was tested using observational data collected during the 1987 Southern California Air Quality Study (SCAQS) and was applied to 25 June (one of the SCAQS episode days); for the Lower Lake Michigan area the procedure was tested using observational data collected during the 1991 Lake Michigan Ozone Study (LMOS) and was applied to 26 June (one of the LMOS episode days).

To provide a basis from which to assess the FDDA simulations, the SAIMM was first exercised for both areas without data assimilation. The SAIMM simulation without FDDA seems to capture many of the important meteorological features of the SCAQS episode day such as the sea breeze and the upslope and downslope flows; however, the observed data are not always well represented in the simulated fields. In particular, the SAIMM simulation indicates westerly flow (outflow from the Los Angeles basin) earlier than observed, and the southerly flow that develops aloft during the evening hours is not simulated. The SAIMM with a zero geostrophic wind and without FDDA is not able to simulate the 26 June meteorology of the Lower Lake Michigan region. While the model generates some physically realistic mesoscale circulation patterns, the prevailing southwesterly flow that is observed over the region during this episode day is not simulated. Due to the large differences between the LMOS no-FDDA simulated fields and the observed data, effective use of the FDDA methodology for this simulation represented a much greater challenge than for the SCAQS simulation.

The nudging-effectiveness experiments were designed to examine the effects of nudging the prognostic wind and temperature variables separately and in combination with one another and to determine (roughly) the optimum nudging coefficients for each. The simulation results for both the Los Angeles and Lower Lake Michigan areas indicate that assimilation of both wind and temperature data provides the best representation of the meteorological fields. The importance of the wind field in air-quality modeling and the indirect benefits derived from the assimilation of the wind data support the use of a larger nudging coefficient for the assimilation of the wind data than for the assimilation of the temperature data. However, strong nudging of the wind components can create some unrealistic airflow patterns over data-sparse regions. In both the SCAQS and LMOS simulations, the best overall simulation results were achieved with a 0.0005 nudging coefficient for assimilation of the wind data and a 0.0001 nudging coefficient for assimilation of the temperature data. Specification of the maximum radii of influence for the interpolation of the data was an important consideration in the simulation of the LMOS episode. Increasing the radii of influence in the objective analysis of the data constrained the simulation over a broader geographical area and contributed to much improved simulation results.

A series of data-reduction experiments, using the SCAQS and LMOS data bases, were performed to investigate the response of SAIMM/FDDA methodology to varying levels of data availability. To accomplish this, monitoring sites were eliminated from the data sets in a series of three site-reduction exercises. A model run was performed with the full data set, and then again after each site reduction to determine how the model adjusted to decreased amounts of observational data. The data-reduction simulation results for both the SCAQS and LMOS episode days indicate that even when the data set is reduced to what might be considered a routine data set, assimilation of the available wind and temperature data provides an improved representation of the meteorological fields when compared with the no-FDDA simulation. As the number of sites is reduced, the simulation errors increase and the effectiveness of the FDDA decreases (i.e., some unusual airflow patterns were simulated over data-sparse or unconstrained subregions of the modeling domain).

The SAIMM/FDDA methodology appears to be a promising technique for the generation of meteorological fields for photochemical modeling. Appropriate specification of the analysis and modeling parameters is essential to successful application of the technique. As these parameters will necessarily be episode- and domain-dependent, thorough testing of the model (including no-FDDA simulation) and evaluation of the simulation results is recommended for each application. Guidelines for evaluation of the meteorological fields will be developed under Phase II of this study. Further study is required and anticipated in order to assess the utility of the SAIMM/FDDA methodology to provide accurate inputs for photochemical modeling.

*U.S. Government Printing Office: 1993 — 750-071/60191

Sharon G. Douglas is with Systems Applications International, San Rafael, CA 94903.
James M. Godowitch and Shao-Hang Chu are the EPA Project Officers (see below).
The complete report, entitled "Application of a Data-Assimilating Prognostic Meteorological Model to Two Urban Areas," (Order No. ;PB93-126571 Cost: $19.50, subject to change) will be available only from;:
National Technical Information Service
5285 Port Royal Road
Springfield, VA 22161
Telephone: 703-487-4650
The EPA Project Officers can be contacted at:
James M. Godowitch
Atmospheric Research and Exposure Assessment Laboratory
U.S. Environmental Protection Agency (MD-80)
Research Triangle Park, NC 27711

Shao-Hang Chu
Office of Air Quality Planning and Standards
U.S. Environmental Protection Agency (MD-14)
Research Triangle Park, NC 27711

United States
Environmental Protection Agency
Center for Environmental Research Information
Cincinnati, OH 45268

BULK RATE
POSTAGE & FEES PAID
EPA
PERMIT No. G-35

Official Business
Penalty for Private Use
$300

EPA/600/SR-92/221